

Minería de datos sobre *streams* de redes sociales, una herramienta al servicio de la Bibliotecología



Sonia Jaramillo Valbuena, Sergio Augusto Cardona y Alejandro Fernández

Universidad del Quindío / sjaramillo@uniquindio.edu.co, sergio_cardona@uniquindio.edu.co,

Universidad Nacional de La Plata / alejandro.fernandez@lflia.info.unlp.edu.ar

Resumen

Los *sistemas* de soporte al trabajo colaborativo son herramientas valiosas en contextos en los cuales se requiere la participación de un grupo de personas para llevar a cabo una determinada tarea. Uno de estos contextos es la Bibliotecología, Archivística y Documentación. Las interacciones entre los usuarios y profesionales de esta área, mediante el uso de herramientas tales como Twitter, Facebook, fuentes RSS y blogs, generan grandes flujos de datos (*streams*) no estructurados. Estos *streams* pueden ser estudiados para analizar aspectos tales como influencia, relaciones de cercanía, opinión y generación de recomendaciones, logrando de esta forma que las bibliotecas obtengan otros beneficios del uso de las tecnologías de información y comunicación. Desde la perspectiva de la minería de datos, el procesamiento de estos *streams* plantea importantes desafíos. El uso de algoritmos de minería en este contexto, implica que en estos se deban considerar aspectos como la adaptación a la alta velocidad en que llegan los datos, la diversidad de las fuentes de datos y su estructura, la variabilidad de los datos en el tiempo y el trabajo sin restricciones de memoria.

Este artículo revisa el estado del arte en lo referente a algoritmos de minería de datos sobre *streams* originados en redes sociales, específicamente, Facebook y Twitter. Se presenta una revisión de las técnicas más representativas de clasificación y agrupamiento, y de cómo cada una de ellas aporta al descubrimiento de conocimiento en el área de la Bibliotecología. Para concluir se presentan algunos de los problemas que son objeto de investigación activa.

Palabras Clave

Minería de flujos de datos
Clasificación
Clustering
Cambio de concepto
Sistemas de soporte al trabajo colaborativo.

Abstract

Data Mining Streams of Social Networks, A Tool to Improve The Library Services. The Groupware systems are a valuable source for disseminating information in contexts in which the participation of a group of people is required to perform a task. One such context is the Library, Archives and Documentation. The interactions among users and professionals in this area, who use tools such as Twitter, Facebook, RSS feeds and blogs, generate a large amount of unstructured data streams. They can be used to the problem of mining topic-specific influence, graph mining, opinion

Keywords

Data stream mining
Classification
Clustering
Concept-drift
CSCW

mining and recommender systems, thus achieving that libraries can obtain maximum benefit from the use of Information and Communication Technologies. From the perspective of data stream mining, the processing of these streams poses significant challenges. The algorithms must be adapted to problems such as: high arrival rate, memory requirements without restrictions, diverse sources of data and concept-drift. In this work, we explore the current state-of-the-art solutions of data stream mining originating from social networks, specifically, Facebook and Twitter. We present a review of the most representative algorithms and how they contribute to knowledge discovery in the area of librarianship. We conclude by presenting some of the problems that are the subject of active research.

Artículo recibido: 05-11-2014. Aceptado: 20-04-2015

Introducción

En la actualidad cientos de aplicaciones generan grandes volúmenes de información en forma de *streams* de datos. Un *stream* es una secuencia continua, potencialmente infinita, de datos que llegan en determinados instantes de tiempo a un sistema para su almacenamiento o procesamiento (Chaudhry, Show y Abdelguerfi, 2005). Ejemplos de *streams* de datos son los provenientes de sensores, sistemas de vigilancia, transacciones bancarias, procesos de *marketing*, mercados de valores y *sistemas Groupware*. Estos últimos tienen como objetivo fundamental soportar a grupos de personas que trabajan para lograr una tarea común y proveen una interfaz para el ambiente compartido (Ellis, Gibbs y Rein, 1991) (Henríquez; Fandos y Gisbert, 2000). Se clasifican a nivel de las aplicaciones en: Sistemas de mensajes, Editores multiusuario, *White Boards*, sistemas de apoyo a las decisiones grupales y salas de reuniones electrónicas, conferencias por computador, agentes inteligentes colaborativos y juegos colaborativos. Los sistemas *Groupware* generan datos no estructurados resultantes de interacciones entre las personas, como es el caso del correo electrónico, herramientas de *control de cambios* en wikis, editores multiusuarios y las redes sociales.

La aplicación de técnicas de minería de datos sobre los datos provenientes de redes sociales puede revelar patrones sobre los individuos inmersos en el ambiente compartido y producir conocimiento que antes no era factible encontrar debido a la variedad y complejidad de la información, como por ejemplo: estrategias de inteligencia de negocios, modelos de predicción avanzada y propagación de historias. Se han desarrollado varios algoritmos para funcionar en ambiente *stream*, y que son aplicables al análisis de información proveniente de redes sociales, entre ellos los algoritmos de clasificación y métodos de segmentación (*clustering*).

El presente trabajo analiza dos importantes redes sociales al servicio de la Bibliotecología, a saber: Facebook y Twitter, y diferentes algoritmos de minería de datos que pueden utilizarse para obtener conocimiento a partir de la información de dichas redes.

El artículo se estructura de la siguiente forma: en primer lugar se presentan aspectos conceptuales sobre las redes sociales Facebook y Twitter, luego se muestran las principales aproximaciones existentes para minería de datos sobre *streams* de redes sociales. En las secciones siguientes se indica la importancia de dichos algoritmos para procesar información proveniente de las redes sociales Facebook y Twitter en beneficio de la Bibliotecología y las ventajas de su aplicación para la obtención de conocimiento al interior del grupo de trabajo. Finalmente se presentan las conclusiones del trabajo.

Facebook y Twitter, dos redes sociales al servicio de la Bibliotecología

Facebook es una red social que permite conectar a una persona con sus amigos y otras personas relacionadas con su entorno (Zuckerberg, 2004). Facebook ofrece servicios tales como la creación de grupos, mensajería, video llamadas, creación de páginas, comunidades de amigos, listado de intereses y manejo de eventos. En la actualidad esta herramienta está siendo utilizada en el ámbito de la Bibliotecología para generar espacios de intercambio, mediante los cuales se pueda mejorar y ampliar la cantidad de servicios de información ofrecidos por las bibliotecas. Entre los principales usos que usuarios y profesionales del área le dan a Facebook está: la generación de perfiles para las bibliotecas, la creación de páginas como modelo de presencia (para difundir los servicios de las bibliotecas) y la formación de grupos (Margaix-Arnal, 2008) (Uribe Tirado y Echavarría Ramírez, 2008).

En el trabajo realizado por Uribe Tirado y Echavarría Ramírez (2008) se presenta un informe con la cantidad de Grupos en Facebook relacionados con el área de la Bibliotecología en Iberoamérica. La consulta realizada consistió en buscar dentro de Facebook las palabras claves: Ciencia de la Información, Bibliotecología, Biblioteconomía, Documentación y Archivística. Allí seleccionaron los grupos cuyo nombre contenía la palabra clave y que eran pertinentes con la búsqueda. Para la búsqueda no se tuvieron en consideración nombres amplios, con más de dos de las palabras claves. Los resultados obtenidos se muestran en la Tabla 1.

Palabra clave	Grupos
Bibliotecología	28
Documentación	145
Archivística	16
Ciencia de la Información	77
Biblioteconomía	11

Tabla 1. Consolidado de grupos relacionados con la Bibliotecología, Documentación y Archivística a septiembre de 2008 (Uribe Tirado y Echavarría Ramírez, 2008)

Para identificar si la tendencia de uso de Facebook en el área de la Bibliotecología había aumentado, a abril de 2015, se realizó esta misma consulta, obteniendo los resultados mostrados en la Tabla 2.

Palabra clave	Grupos
Bibliotecología	127
Documentación	240
Archivística	62
Ciencia de la Información	120
Biblioteconomía	108

Tabla 2. Consolidado de grupos relacionados con la Bibliotecología, Documentación y Archivística a abril de 2015

Los resultados evidencian que la creación de grupos en Facebook para manejar aspectos referentes al área de Bibliotecología, Archivística y Documentación, crece con el paso de los años, lo cual tiene coherencia con la tendencia impuesta por la Web 2.0 de compartir información.

Por su parte, Twitter (Dorsey; Glass; Williams y Stone, 2006) es una red que permite conectarse con amigos y obtener en tiempo real información de eventos y temas de interés. Un tuit es una publicación o mensaje en esta plataforma de micro blog o una

actualización en el estado de un usuario. Los tuits pueden acompañarse de etiquetas que permitan precisar el enfoque de la publicación, o crear una relación con un tema de conversación que es tendencia. Un tuit puede aparecer destacado en la cuenta de un usuario debido a la trascendencia que este haya tenido entre la comunidad de seguidores. Twitter ofrece la posibilidad de fijar tuits para darle una mayor visibilidad a un mensaje dentro de la línea de tiempo y determinar los tuits que son temas relevantes en el momento, dependiendo del interés generado o por la importancia del tema dentro de la comunidad (Seven, 2015).

Twitter puede ser utilizado por el bibliotecario para ofrecer, en tiempo real, servicios de referencias e información, proporcionando a cada usuario la dirección web de los servicios ofrecidos (Peña, 2013). De igual forma, en el área de la Bibliotecología, es posible su utilización para determinar por ejemplo, temáticas y libros que son tendencia a nivel mundial. Esta información podría aplicarse, entre otros fines, para mejorar el proceso de adquisición y reserva de material bibliográfico.

Procesamiento y análisis de *streams* de datos

Para procesar *streams* de datos existen dos modelos de cómputo. El primero de ellos, el evolutivo, que toma en consideración el histórico de los datos. El segundo, llamado modelo de ventana, se aplica a sistemas en los cuales se requiere considerar los datos recibidos más recientemente (Callau Zori, 2013). Este modelo funciona de forma similar a una “cola”, ya que cuando se inserta información al final de la ventana, la transacción que se encuentra en el extremo inicial se saca de la cola. En este modelo, solo el *sub-stream* acotado por las dimensiones de la ventana se procesa y almacena. La ventana tiene un tamaño y un desplazamiento, ver Figura 1. El tamaño de la ventana puede ser fijo o no, dependiendo de ello los límites de la ventana se desplazan. Cuando el tamaño de la ventana es fijo y la ventana se desliza, los puntos extremos se mueven, a ambos lados en la misma cantidad a lo largo del *stream* de datos.

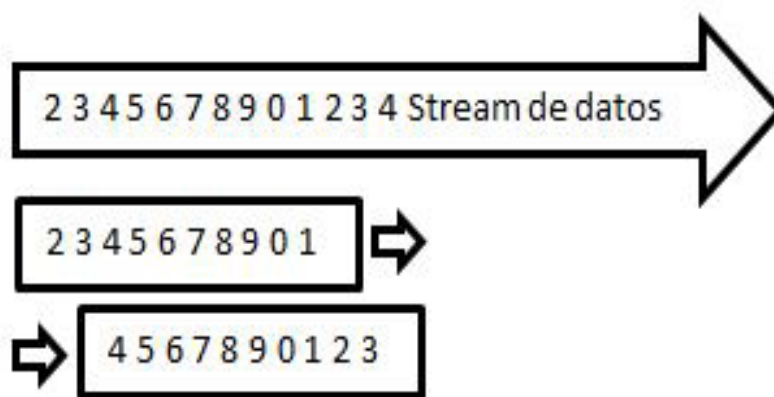


Figura 1. Ventana deslizante de tamaño 10 y desplazamiento 2 (Finlay, Pears y Connor, 2014)

La obtención de conocimiento a partir de los *streams*, requiere la aplicación de técnicas de minería de datos. Se entiende por minería de datos, al proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos (Fayyad; Piatetsky Shapiro y Smyth, 1996).

Existen diversos marcos de trabajo (*frameworks*) de minería sobre *streams* de datos, que hacen uso de los modelos evolutivo y de ventana. Algunos de los más relevantes

para minería de datos sobre *streams*, son *Massive Online Analysis* (MOA) y *Scalable Advanced Massive Online Analysis* (SAMOA), ambos open-source (University of Waikato, 2014). MOA incluye, entre otros, la implementación de algoritmos de clasificación y *clustering*. La Tabla 3 presenta algunos de los algoritmos implementados en MOA.

Técnica	Descripción	Algoritmos implementados en el Framework MOA
Clasificación	Permite asignar una categoría a ejemplos antes no vistos. Por ejemplo, conociendo las palabras claves de la introducción de un libro o artículo es posible determinar la temática con la cual está relacionado.	Clasificadores Bayesianos
		Árboles de decisión de Hoeffding
		Meta clasificadores (combinación de modelos cuyas predicción individuales se combinan de alguna forma para formar una predicción final)
Clustering	Se utiliza para particionar los datos en grupos de objetos similares. Una colección de objetos con características similares entre sí, recibe el nombre de cluster. El bibliotecario podría tener <i>clusters</i> , con libros correspondientes a determinadas áreas temáticas. Sobre estos <i>clusters</i> podría, entre otras funciones, calcular estadísticas, identificar libros pertenecientes a cada <i>cluster</i> . También sería posible identificar áreas temáticas afines, calculando medidas de cercanía entre los <i>cluster</i> .	StreamKM++
		CluStream
		ClusTree
		DenStream
		D-Stream
		CobWeb

Tabla 3. Algoritmos de clasificación y *clustering* implementados en MOA

Por su parte, SAMOA permite realizar minería sobre *streams* de datos en un ambiente *cluster* o nube. Mediante este *framework* es posible desarrollar un algoritmo de Machine Learning distribuido y ejecutarlo sobre múltiples motores de procesamiento de *streaming* (SPE), tales como S4 (Apache Software Foundation, 2013) o Storm (Apache Software Foundation, 2014). SAMOA permite hacer uso de los algoritmos implementados en MOA y adicionalmente brinda algoritmos como los presentados en la Tabla 4.

Técnica	Algoritmos implementados en el Framework MOA
Clasificación	Árbol de decisión vertical de Hoeffding (VHT), metaclasificadores (Bagging, Adaptive Bagging, Boosting)
Clustering	CluStream distribuido

Tabla 4. Algoritmos de clasificación y *clustering* implementados en SAMOA

Trabajos previos en los que se aplican técnicas de minería de datos sobre *streams* provenientes de redes sociales

Con relación al uso de la minería de datos para extraer conocimiento de las redes sociales se han dado varias aproximaciones. En Bifet y Frank (2010) se aplican técnicas de clasificación para minería de opinión y análisis de sentimientos sobre Twitter (Twitter, Inc., 2014). El análisis de opinión permite detectar información obtenida de un texto y clasificar la orientación expresada en dichos texto. Básicamente, lo que se pretende con el análisis de opiniones es comprender la actitud que tiene la persona que escribe sobre una temática o la polaridad contextual (positiva, negativa o neutra) que expresa. Entre la información que puede obtenerse empleando análisis de opiniones se tiene: nivel de fidelización de clientes, polaridad (en lo referente a críticas sobre productos, servicios, entre otros), tendencias de mercado y predicciones sobre candidatos que aspiran a ser elegidos. Al respecto, se resalta que las opiniones de las personas expresadas en lenguaje natural son un importante recurso para la construcción de perfiles de usuario en la gestión de los Sistemas de recomendación.

En Balasubramanyan; Routledge y Smith (2010) se comparan medidas de opinión pública obtenidas mediante encuestas con medidas de sentimiento calculadas desde texto. Se encuentra que lo expresado en las encuestas se correlaciona con las frecuencias de palabras de opinión en los mensajes analizados, esto les permite identificar tendencias a gran escala. Por su parte, en Aston; Munson; Liddle; Hartshaw y Livingston (2014) se presenta el algoritmo Winnow modificado para análisis de opinión en línea sobre redes sociales. Este algoritmo, que es utilizado para el aprendizaje de un clasificador lineal a partir de ejemplos pre-etiquetados, se integra a MOA con el fin de entrenar y correr el clasificador. En AT&T Labs (2013) se hace referencia a un motor de búsqueda visual online que realiza minería de texto en tiempo real. Los tuits recuperados son automáticamente categorizados y resumidos haciendo uso de técnicas de análisis semántico y métodos de *clustering*. Por su parte, Wood; Zheludev y Treleaven (2014) presentan a UCL's SocialSTORM, una plataforma para minar datos provenientes de Twitter, Facebook, fuentes RSS y blogs. La plataforma permite monitorear datos de *streams* históricos y actuales. Fue diseñada para aplicaciones comerciales, pero puede ser usada en otros ámbitos, tales como supervisión de seguridad y monitoreo de variables macroeconómicas. En Popovici, Weiler y Grossniklaus (2014) extienden el algoritmo de *clustering* DenStream y lo aplican sobre datos provenientes de redes sociales para lograr detección de tópicos en tiempo real. Finalmente, en Bifet; Holmes; Pfahringer y Gavalda (2011) se presenta a MOA-TweetReader, un sistema de tiempo real que lee tuits, detecta cambios y halla los términos cuya frecuencia ha cambiado.

Ahora bien, la aplicación de técnicas de minería de datos en bibliotecas, *bibliomining*, es analizada en varios trabajos. Doszkocs (1991) profundiza en el uso de redes neuronales para identificar asociaciones entre los documentos. Por su parte, Chau (2000) plantea cómo la minería de datos Web beneficia la personalización de los servicios de referencia electrónica. Nicholson (2005) menciona el uso de la minería de datos para brindar un mejor acceso a la colección. Finalmente, Candás Romero (2006) y Herrera Varela (2006) enuncian aplicaciones prácticas de la minería de datos en beneficio de los servicios de las bibliotecas 2.0. Entre las aplicaciones expuestas se destacan: evaluación bibliotecaria, selección, administración y adquisición de fondos, apoyo a la toma de decisiones, organización de la colección, presentación de resultados, catalogación automatizada partiendo del títulos de libros y revistas, descubrimiento de información oculta en colecciones y sistemas de recomendaciones. Las referencias anteriores se enfocan en la aplicación de técnicas de minería en ambientes por lotes (modo *batch*). Existen numerosas referencias sobre entornos *batch* pero se omiten por no corresponder al tema del presente artículo. Los sistemas tipo *batch* pueden verse como el extremo opuesto al procesamiento de *streams* de datos. En las aplicaciones

de tipo *stream* los datos se generan en tiempo real y su análisis debe ejecutarse sobre la marcha, esto quiere decir, que requieren, durante su procesamiento, respuesta con unos plazos de tiempo acotados. Además, la tasa de llegada y el volumen de los datos es tan grande que es impráctico su almacenamiento. Las limitaciones inherentes a los *streams* de datos en términos de tiempo y memoria hacen que algoritmos de clasificación y *clustering* tradicionales no puedan ser directamente aplicados, y por ende, deban adaptarse para considerar la evolución de los datos en el tiempo y trabajar con datos en tan solo una pasada, es decir, luego de que el dato llega y se procesa, éste generalmente caduca y no se tiene disponible para análisis posteriores (Jiang y Gruenwald, 2006) (Domingos y Hulten, 2001).

El uso de minería sobre *streams* de datos provenientes de redes sociales, en el área de la Bibliotecología, es un tema relativamente nuevo. Se destaca el proyecto realizado por la Biblioteca del Congreso de los Estados Unidos, que consiste en procesar los comentarios efectuados a través de Twitter, mediante la aplicación técnicas de minería de datos, para extraer conocimiento aplicable, entre otros aspectos, por la protección de los derechos de autor (BiblioInstruccion, 2010).

La Tabla 5 presenta una clasificación de las aproximaciones presentadas previamente con base en el modelo de cómputo y los algoritmos de extracción de conocimiento.

Aplicación de técnicas de minería de datos en información proveniente de redes sociales utilizadas en el área de la Bibliotecología

La aplicación de técnicas de minería de datos sobre redes sociales puede hacerse principalmente desde dos contextos. El primero de ellos, el análisis de vínculos o minería de grafos, tiene como objetivo medir la dinámica de la población, descubrir patrones referentes a la formación de la comunidad y analizar la forma de dispersión de la información dentro de la misma. El segundo enfoque, en el que se analiza el contenido de los mensajes, considera realizar clasificación y agrupamiento, análisis de opinión, detección de temas e identificación de identidades. Algunas de las dificultades que se plantean para la aplicación de técnicas de aprendizaje sobre estos flujos de información son la cantidad de ruido, el flujo permanente y la longitud corta de los textos.

La aplicación de minería de información en redes sociales trae importantes ventajas, al permitir obtener conocimiento en tiempo real a partir de texto, consolidándose en un instrumento potencial de análisis y permitiendo establecer correlaciones entre los textos y las noticias de medios o *mainstream* (Anguiano Hernández, 2011).

La minería sobre datos provenientes de Facebook y Twitter, puede beneficiar a la Bibliotecología en los siguientes aspectos:

- » Uso de sistemas de recomendación basados en confianza para sugerir elementos que se ajustan a las necesidades del usuario de la biblioteca partiendo del conocimiento de las preferencias del usuario (micro segmentación) y el contexto en que se desenvuelve.
- » Análisis de las consultas y comportamiento de los usuarios en el OPAC.
- » Inferir opiniones tanto a nivel particular como grupal sobre determinados asuntos o temáticas relacionados con la Bibliotecología, por ejemplo en lo referente a calidad de los servicios prestados por la biblioteca.
- » Determinación de las preferencias de los usuarios de la biblioteca.

Referencia	Contexto	Dataset
(Bifet y Frank, 2010)	Análisis de sentimientos de <i>streams</i> provenientes de Twitter.	-twittersentiment.appspot.com -Edinburgh Corpora
(Balasubramanyan; Routledge y Smith, 2010)	Correlacionan las observaciones expresadas en encuestas de opinión pública con los resultados del análisis de los sentimientos efectuado sobre textos de Twitter	1 billón de tuits publicados en los años 2008 y 2009
(Aston; Munson; Liddle; Hartshaw y Livingston, 2014)	Análisis de sentimientos. Se evalúa la exactitud del clasificador	-Sanders Corpus (5513 tweets clasificados manualmente) -STS (2034 tuits) -Gold SentiStrength (4242 mensajes provenientes de MySpace, YouTube, Digg, BBC, Runners World)
(AT&T Labs, 2013)	Un motor de búsqueda visual en línea para minería de texto sobre <i>streams</i> en tiempo real	Se obtienen los datos mediante el Twitter's <i>streaming</i> API, de acuerdo a la palabra clave seleccionada por el usuario. Los tópicos disponibles son deportes, visualización, datos, ciencia, tecnología y TED (Tecnología, Entretenimiento, Diseño)
(Wood; Zheludev y Treleven, 2014)	Minería sobre datos sociales, permite identificación de tendencias de cambio, sentimientos globales y propagación de historias	Los datos se obtienen mediante el Twitter API y el Facebook's Graph API
(Popovici; Weiler y Grossniklaus, 2014)	<i>Clustering on-line</i> para detección de tópicos en tiempo real en redes sociales	SNOW challenge (1,041,227 tuits)
(Bifet; Holmes; Pfahringer y Gavalda, 2011)	Detección de cambios de sentimiento en <i>streams</i> de tuits.	Obtenidos mediante APIS: Twitter Application Programming Interface (API) y REST APIs.
(BibliInstruccion, 2010)	Identificación de tendencias de cambio, análisis de sentimientos, categorización	Twitter

Tabla 5. Clasificación de algunos de los trabajos realizados sobre minería de datos en redes sociales

- » Obtener datos demográficos importantes (género, idioma o localización) sobre los usuarios.
- » Segmentar usuarios con gran precisión; esta información puede ser la base de una estrategia de publicidad de los servicios de la biblioteca.
- » Generación de gráficos sociales que sinteticen la información de los usuarios, o de los profesionales del área de la Bibliotecología.
- » Predecir cuestiones indirectas, por ejemplo, correlaciones estadísticas directas entre el estado de ánimo de los usuarios y la tasa de préstamos de libros.
- » Evaluación de resultados y apoyo a toma de decisiones, para mejorar futuras actuaciones.

- » Identificar tendencias globales, por ejemplo libros que son ampliamente leídos a nivel mundial.
- » Analizar formas de interacción que llevan a un comportamiento colectivo complejo; determinar puntos atípicos y valores extremos dentro de la red de usuarios de la biblioteca.
- » Obtener información actualizada sobre los últimos mensajes escritos en Twitter referente a determinada temática o publicación.
- » Predecir cuáles serán las necesidades de los usuarios de la biblioteca.
- » Identificar aspectos relacionados por similitud, derivación o causa/efecto (Candás Romero, 2006).
- » Resumen automático de documentos.
- » Descripción completa y adecuada de los ítems.
- » Optimizar los servicios ofrecidos (por ejemplo respecto a la reorganización de contenidos).
- » Personalización, mediante la identificación de temas de interés, contribuyendo a mejorar los sistemas de Difusión Selectiva de Información, DSI.
- » Determinar cuáles materiales tienen baja tasa de consulta porque han sido mal clasificados.
- » Toma de decisiones respecto a acuerdos inter-bibliotecarios, por ejemplo analizando costo del préstamo frente a costo de compra del material.

Conclusiones

La Bibliotecología, Archivística y Documentación es un área propicia para la aplicación de técnicas de minería de datos. Allí se cuenta con una gran cantidad de información, que puede ser procesada para mejorar el potencial de la biblioteca: ahorrar dinero, resolver de forma eficiente las necesidades del usuario, identificar problemas dentro de la colección, seleccionar programas y justificar la toma de decisiones. En general, esto puede verse como gerencia del conocimiento. Hasta ahora, el *bibliomining* ha sido usado efectivamente en entornos tipo *batch* (por lotes), pero su aplicación sobre *streams* de datos es un tema relativamente nuevo e inexplorado.

A pesar de lo novedoso del tema, puede afirmarse que el uso de técnicas de minería sobre *streams* de datos provenientes de Facebook y Twitter beneficia el entorno bibliotecológico en gran medida. Es una importante herramienta para obtener conocimiento en “tiempo real” y predecir patrones. Su uso hace posible, entre otros objetivos, identificar fortalezas y/o debilidades mediante el análisis de opiniones que puedan considerarse como marcadamente positivas o negativas, para con base en ello, poner prácticas acciones correctivas al interior del grupo de trabajo y contribuir a que se dé un proceso real de mejoramiento de la calidad de los procesos al interior de la biblioteca.

Cualquiera que sea el enfoque seguido para la aplicación de minería de datos sobre *streams* (minería de grafos o análisis de contenido de los mensajes), es necesario tener en consideración las limitaciones inherentes a los flujos de datos en lo referente a tiempo, memoria y cambios en distribuciones subyacentes. En la actualidad, estas limitantes son área de investigación activa; al igual que lo es el manejo de datos no estructurados. Respecto a estos últimos, la Web semántica se proyecta como una importante alternativa para mejorar los resultados obtenidos mediante la aplicación de técnicas de Aprendizaje Automático, esta combinación puede ser aplicable para extraer recomendaciones basadas en ontologías.

Referencias bibliográficas

- » Anguiano Hernández, E. 2011. *Introducción a la Minería de Información en Redes Sociales*. <<http://ccc.inaoep.mx/labtl/uploads/Main/intro-mineria-redes-sociales.pdf>>. [Consulta: 20 abril 2015].
- » Apache Software Foundation. 2013. *S4 distributed stream computing platform*. <<http://incubator.apache.org/s4/>>. [Consulta: 20 abril 2015].
- » Apache Software Foundation. 2014. *Storm, distributed and fault-tolerant realtime computation*. <<https://storm.incubator.apache.org/>>. [Consulta: 20 abril 2015].
- » Aston, N.; T. Munson; J. Liddle; G. Hartshaw y D. Livingston. 2014. Sentiment Analysis on the Social Networks Using Stream Algorithms. En *Journal of Data Analysis and Information Processing*. Vol. 2, 60-66.
- » AT&T Labs. 2013. *An Online Visual Search Engine for Mining Streaming Text Data in Real-Time*. <http://www.research.att.com/techdocs/TD_101174.pdf> [Consulta: 20 abril 2015].
- » Balasubramanyan, Ramnath; Brendan O'Connor; Bryan R. Routledge y Noah A. Smith. 2010. From Tweets to Polls : Linking Text Sentiment to Public Opinion Time Series. En *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. Association for the Advancement of Artificial Intelligence. <<https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1536/1842>> [Consulta: 20 abril 2015].
- » Biblioinstrucción. 2010. *Tweeter, las bibliotecas y el almacenaje del conocimiento de la Web 2.0*. <<http://biblioinstruccion.blogspot.com/2010/10/tweeter-las-bibliotecas-y-el-almacenaje.html>> [Consulta: 20 abril 2015].
- » Bifet, A. y R. Kirkby, 2009. *Massive Online Analysis Manual*. <<http://www.cs.waikato.ac.nz/~abifet/MOA/Manual.pdf>>. [Consulta: 20 abril 2015].
- » Bifet, A.; G. Holmes; B. Pfahringer; R. Kirkby y R. Gavalda. 2009. New Ensemble Methods for Evolving Data Streams. En *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM. p. 139-148.
- » Bifet, Albert y E. Frank. 2010. *Sentiment Knowledge Discovery in Twitter Streaming Data*. <<http://www.cs.waikato.ac.nz/~ml/publications/2010/Twitter-crc.pdf>>. [Consulta: 20 abril 2015].
- » Bifet, Albert; Geoffrey Holmes; Bernhard Pfahringer y Ricard Gavalda. 2011. Detecting Sentiment Change in Twitter Streaming Data. En Diethe, Tom; José L. Balcázar; John Shawe-Taylor y Cristina Tîrnuca, ed. *Proceedings of the Second Workshop on Applications of Pattern Analysis* (2º: 19-21 Octubre 2011: Castro Urdiales, España.) WAPA. JMLR Proceedings 17, p. 5-11. JMLR.org. <<http://dblp.uni-trier.de/db/journals/jmlr/jmlr17.html#BifetHPG11>>. [Consulta: 20 abril 2015].
- » Callau Zori, M. 2013. *Algoritmos de agrupación para flujos de datos en entornos centralizados y distribuidos*. <<http://oa.upm.es/22467/>>. [Consulta: 20 abril 2015].
- » Candás Romero, J. 2006. *Minería de datos en bibliotecas: bibliominería*. <<http://bid.ub.edu/17candaz.htm>>. [Consulta: 20 abril 2015].

- » Chau, M. 2000. *Mediating off-site electronic reference services : human-computer interactions between librarians and web mining technology*. <<http://osulibrary.oregonstate.edu/mediating-site-electronic-reference-services-human-computer-interactions-between-librarians-and-web>>. [Consulta: 20 abril 2015].
- » Chaudhry, Nauman; Kevin Shaw y Mahdi Abdelguerfi. 2005. *Stream data management*. Nueva York: Springer. Kluwer international series on advances in database system. Vol. 30.
- » Domingos, Pedro y Geoff Hulten. 2001. *Catching up with the data: Research issues in mining data streams*. Trabajo presentado en Workshop on Research Issues in Data Mining and Knowledge Discovery, Santa Barbara, CA el 20 de mayo de 2001.
- » Dorsey, J.; N. Glass; E. Williams y B. Stone. 2006. *Twitter*. <<http://twitter.com>>. [Consulta: 20 abril 2015].
- » Doszko, T. 1991. *Neural networks in libraries*. <<http://web.simmons.edu/~chen/nit/NIT%2791/027-dos.htm>>. [Consulta: 20 abril 2015].
- » Ellis, Clarence A.; Simon J. Gibbs y Gail Rein. 1991. Groupware: some issues and experiences. En: *Communications of the ACM*. Vol. 34, no. 1, 39-58. <<http://doi.acm.org/10.1145/99977.99987>>. [Consulta: 20 abril 2015].
- » Fayyad, Usama; Gregory Piatetsky Shapiro y Padhraic Smyth. 1996. From Data Mining to Knowledge Discovery in Databases. En *AI Magazine*. Vol. 17, no. 3, 37-54. <<http://www.aaai.org/ojs/index.php/aimagazine/article/view/1230>> [Consulta: 20 abril 2015].
- » Finlay, J.; R. Pears y A. M. Connor. 2014. *Data stream mining for predicting software build outcomes*. <http://aut.researchgateway.ac.nz/bitstream/handle/10292/7086/IST_DataStreamMining.pdf?sequence=5>. [Consulta: 20 abril 2015].
- » Henríquez, P.; M. Fandos, y M. Gisbert. 2000. La comunicación asíncrona en un curso del Proyecto Intercampus. En Pérez, R., coord. *Redes, multimedia y diseños virtuales*. Oviedo: Departamento de Ciencias de la Educación, p. 269-276.
- » Herrera Varela, R. 2006. *Bibliomining: minería de datos y descubrimiento de conocimiento en bases de datos aplicados al ámbito bibliotecario*. <http://bibliotecarios.cl/conferencia_2006/C2006_019.pdf>. [Consulta: 20 abril 2015].
- » Jiang, Nan y Le Gruenwald. 2006. Research issues in data stream association rule mining. En *SIGMOD Record*. Vol. 35, no. 1, 14-19.
- » Margaix-Arnal, D. 2008. Las bibliotecas universitarias y Facebook: cómo y por qué estar presentes. En *El profesional de la información*. Vol. 17, no. 6, 589-601.
- » Nicholson, S. 2005. *The Basis for Bibliomining: Frameworks for Bringing Together Usage-Based Data Mining and Bibliometrics through Data Warehousing in Digital*. <<http://bibliomining.com/nicholson/nicholsonbibliointro.html>>. [Consulta: 20 abril 2015].
- » Peña, L. M. 2013. *El bibliotecario web 2.0*. <<http://bibliotecahoy.blogspot.com/2013/10/el-bibliotecario-web-20.html>>. [Consulta: 20 abril 2015].
- » Popovici, Robert; Andreas Weiler y Michael Grossniklaus. 2014. On-line Clustering for Real-Time Topic Detection in Social Media Streaming Data. En Papadopoulos, S.; D. Corney y L. Aiello, eds. *Proceedings of the SNOW 2014 Data Challenge*, (8 abril 2014: Seoul, Korea). <<http://ceur-ws.org/Vol-1150/popovici.pdf>> [Consulta: 20 abril 2015].

- » Seven, G. 2015. *Significado de Tweet*. <<http://www.significados.com/tweet/>>. [Consulta: 20 abril 2015].
- » Twitter, Inc. 2014. *About Twitter*. <<https://twitter.com/>>. [Consulta: 20 abril 2015].
- » University of Waikato. 2014. *MOA Massive Online Analysis*. <<http://moa.cms.waikato.ac.nz/>>. [Consulta: 20 abril 2015].
- » Uribe Tirado, Alejandro y Andrés Felipe Echavarría Ramírez. 2008. Facebook como red de profesionales de bibliotecología documentación y archivística en Iberoamérica. En *El profesional de la información*. Vol. 17, no. 6, 670-676. <<http://www.elprofesionaldelainformacion.com/contenidos/2008/noviembre/11.pdf>> [Consulta: 20 abril 2015].
- » Wood, R.; I. Zheludev y P. Treleaven. 2014. *Mining Social Data with UCL's SocialSTORM Platform*. <https://www.academia.edu/4144373/Mining_Social_Data_with_UCLs_Social_STORM_Platform>. [Consulta: 20 abril 2015].
- » Zuckerberg, M. 2004. *Facebook*. <https://www.facebook.com/?_rdr>. [Consulta: 20 abril 2015].